



## PATENT ABSTRACTS OF JAPAN

(11) Publication number: **05094199 A**(43) Date of publication of application: **16.04.93**

(51) Int. Cl. **G10L 5/04**  
**G10L 3/00**  
**G10L 9/14**

(21) Application number: **03253863**(22) Date of filing: **01.10.91**(71) Applicant: **SANYO ELECTRIC CO LTD**

(72) Inventor: **KITAMURA TORU**  
**FUJIMOTO MITSUO**

(54) **RESIDUAL DRIVING TYPE SPEECH  
 SYNTHESIZING DEVICE**

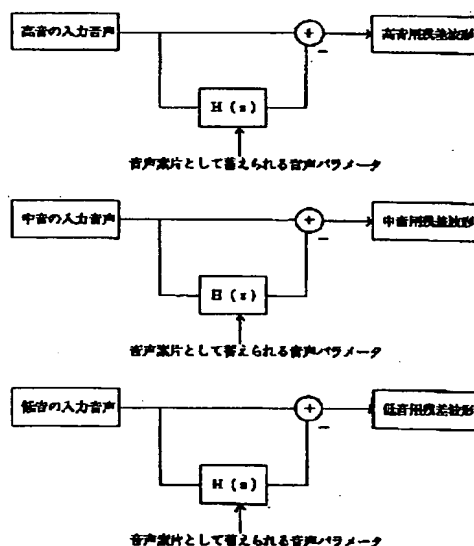
## (57) Abstract:

**PURPOSE:** To obtain a composite voice of a more natural and high quality by constituting the device so that even if a residual waveform of any pitch period of a high-pitched sound, a middle-pitched sound and a low-pitched sound is used to a voice element piece selected in order to synthesize a voice, waveform data of all residual signals that can be selected is extracted by a reverse filter in which 2 voice parameter itself provided to actual synthesis is a coefficient, and a sound source and the parameter are extracted from the same voice in spite of a change of the pitch.

**CONSTITUTION:** A voice parameter of a voice element piece unit of a voice element piece memory is used as a coefficient of a voice analytic filter consisting of a reverse filter of a voice synthesizing filter, waveform data of residual signals of a voice of three kinds of different pitch periods obtained by inputting a voice of pitch periods of a high-pitched sound, a middle-pitched sound and a low-pitched sound to this analytic filter, respectively are extracted, and these waveform data are

accumulated in a residual waveform memory.

COPYRIGHT: (C)1993,JPO&amp;Japio



JP A05-094199

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平5-94199

(43) 公開日 平成5年(1993)4月16日

(51) Int. Cl. <sup>5</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G10L 5/04		F 8946-5H		
3/00		H 8946-5H		
9/14		L 8946-5H		

審査請求 未請求 請求項の数 1 (全10頁)

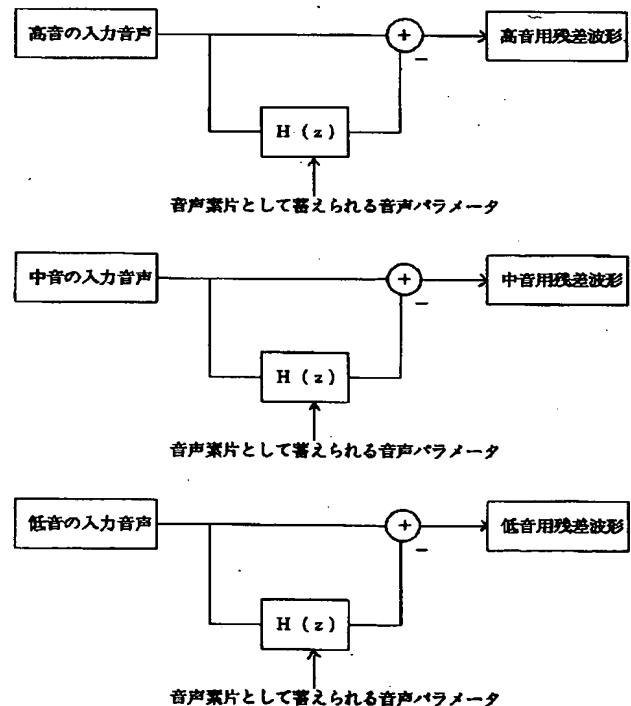
(21) 出願番号	特願平3-253863	(71) 出願人	000001889 三洋電機株式会社 大阪府守口市京阪本通 2 丁目18番地
(22) 出願日	平成3年(1991)10月1日	(72) 発明者	北村 徹 守口市京阪本通 2 丁目18番地 三洋電機株式会社内
		(72) 発明者	藤本 光男 守口市京阪本通 2 丁目18番地 三洋電機株式会社内
		(74) 代理人	弁理士 西野 卓嗣

(54) 【発明の名称】 残差駆動型音声合成装置

(57) 【要約】

【構成】 本発明の残差駆動型音声合成装置は、音声合成フィルタの逆フィルタからなる音声分析フィルタの係数として、音声素片メモリの音声素片単位の音声パラメータを用い、この分析フィルタに高音、中音、低音のピッチ周期の音声をそれぞれ入力することにより得られる異なる3種類のピッチ周期の音声の残差信号の波形データを抽出し、これらの波形データを残差波形メモリに蓄えたものである。

【効果】 本発明によれば、音声を合成するために選択された音声素片に対して、高音、中音、低音のいずれのピッチ周期の残差波形が駆動音源として用いられても、選択され得る全ての残差信号の波形データが、実際の合成に供せられる音声パラメータそのものを係数とした逆フィルタにより抽出されたものであるため、ピッチの変更にかかわらず、音源とパラメータが同一音声から抽出したことになるため、より自然で高品質な合成音声を得られる。



## 【特許請求の範囲】

【請求項 1】 LPC、LSP などの線形予測系の音声特徴パラメータを音声素片単位で蓄えた音声素片メモリ、発声すべき音声の音声素片を示す記号列を生成する音韻記号列生成部、及び該音韻記号列生成部により生成された記号列に基づいて音声素片メモリから読み出した音声素片を順次接続する音声素片接続部を備えた音声特徴パラメータ制御手段、

各音声素片毎に対応したピッチ周期の異なる複数の残差信号の波形データ群を蓄えた残差波形メモリ、発声すべき音声のピッチ周期変化を示したピッチパターンを生成するピッチパターン生成部、及び上記残差波形メモリの音声素片に対応した残差信号の波形データ群の中から該ピッチパターン生成部で決定された各時点でのピッチ周期に応じた残差波形データを選択する残差選択回路を備えた駆動音源制御手段、

該残差選択回路で選択された残差を駆動音源とすると共に、上記該音声素片接続部によって接続された音声素片の音声パラメータを係数として音声を合成する線形予測系の音声合成フィルタを備えた音声合成手段からなる残差駆動型音声合成装置に於て、

上記残差波形メモリに蓄えるべきピッチ周期の異なる複数の残差信号の波形データは、これと対応する上記音声素片メモリの音声素片の音声パラメータを係数として、上記音声合成フィルタの逆フィルタからなる音声分析フィルタに、それぞれ異なるピッチ周期の音声を入力することにより得ることを特徴とした残差駆動型音声合成装置。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】 本発明は、任意の言葉を発声することが可能な規則音声合成装置、特に残差駆動を行う残差駆動型規則音声合成装置に関する。

## 【0002】

【従来の技術】 近年、任意の文章から音声を合成するための規則合成手法の研究が盛んであり、現在では、新聞の校閲装置や盲人用読書機などに試作、実用化されているものがある。

【0003】 任意の文章から音声を合成するための規則合成装置は、例えば、テキスト入力に対し、文章解析を行って読みがなやアクセントを決定し、音韻規則から、必要な合成単位である音声素片（例えば CVC 単位）を決定して結合し、韻律規則から、声の高さなどを決定して、音声パラメータの時系列とピッチパターンを生成し、これらのパラメータから音源とデジタルフィルタを構成することにより、合成音声を生成する。

【0004】 さて、このような音声合成手法に用いる音声パラメータとしては、LPC、LSP などの線形予測系のパラメータやフォルマントなどが一般的であり、一方、音源としては、メモリの削減と処理の簡単化のた

め、インパルスと白色雑音を用いられていた。

【0005】 而して、LPC、LSP などの線形予測系の音声合成手法では、予測残差を駆動音源として用いることにより、原音声に近い合成音声を得られることが知られている。

【0006】 従って、文字等の入力により任意の音声を発声可能な規則合成についても、駆動音源として残差を入力することにより、上記の原理から、高品質な合成音を得られることが期待され、このような残差駆動型の規則音声合成装置が提案されており（特願平 2 - 2 4 9 4 9 3 号）、これについて以下に概説する。

【0007】 図 1 は、既提案の残差駆動型の規則音声合成装置の構成をしたものであり、この装置によると、発声すべき文字列が文字列バッファ（1）に入力されると、音韻記号列生成部（2）は入力された文字列を音韻記号列に変換する。例えば、「た＊べにき＊た」（但し、＊はアクセント位置を示す記号）という文字列が入力されると「tabenikita」という音韻記号列に変換する。

【0008】 次に、選択回路 1（3）は、音韻記号列から必要な音声素片を順次、決定選択し、音声素片メモリ（4）に蓄えられた音声素片のうち必要な音声素片が、音声素片接続部（5）で接続される。音声素片の単位としては、CVC（子音＋母音＋子音）、あるいは、CV（子音＋母音）と VC（母音＋子音）を併用するものなども用いられるが、例として簡単のため、CV（子音＋母音）すなわち音節を単位とするものを考えると、「ta、be、ni、ki、ta」が必要な音声素片として順次選択され接続される。ここで接続された音声素片は、音声パラメータとして音声パラメータバッファ（6）に蓄えられ、係数として合成フィルタ（11）に与えられる。

【0009】 一方、アクセント位置等のイントネーションを表す記号も、発声すべき文字列とともに入力され、文字列バッファ（1）から、ピッチパターン生成部（7）に与えられると、ピッチパターン生成部（7）は、発声文全体のピッチ（音程）を決定する。例えば、「た＊べにき＊た」という入力の場合、「た」と「き」にアクセントが存在するので、第 5 図に示すようなピッチパターンとなる。ピッチパターン生成部（7）では、文全体にピッチが降下するフレーズ成分と、アクセント位置でピッチが高くなるアクセント成分が加えられて、ピッチパターンが生成される。

【0010】 また、残差波形メモリ（8）には、図 3 に示す如く、各音声素片に対応して、駆動音源として利用するための残差波形が蓄えられており、選択回路 1

（3）で決定選択された音声素片に対応して、必要な残差波形が選択される。例の場合、「ta、be、ni、ki、ta」の順で対応する残差波形が選択される。さらに、この残差波形メモリ（8）には、やはり図 3 に示

す如く、各音声素片に対してピッチの異なる複数の残差波形が蓄えられており、ピッチパターン生成部(7)で生成されたピッチに応じて、選択回路2(9)が適切なピッチの残差波形を選択決定し、駆動音源生成部(10)に蓄える。最後に、選択された残差波形は、駆動音源生成部(10)で、ピッチパターン生成部(7)で生成されたピッチに一致する値にピッチ変更が施され、所望のピッチの残差波形が生成される。

【0011】このようにして生成された残差波形は、駆動音源として合成フィルタ(11)に入力され、合成フィルタ(11)で合成音声が生産される。合成音声はD/A変換器(12)を経て、スピーカ(13)から出力される。

【0012】このような既提案装置の駆動音源生成部(10)の動作について、以下にさらに説明を加える。

【0013】まず、図3は残差波形メモリ(8)に蓄えられている残差信号の波形データ例を示したものである。このメモリ(8)の波形データは、CV(子音+母音)構成の音節を音声素片の単位とした場合に対応しており、各音声素片に対応して、それぞれピッチ周期が異なる3形態、即ち、高音用残差波形(ピッチ周期の短い残差)、中音用残差波形(ピッチ周期の中程度残差)、低音用残差波形(ピッチ周期の長い残差)が蓄えられている。そして、このような残差波形メモリ(8)に蓄えられている残差信号の波形データは、図4の駆動音源生成部(10)によって、以下の如く処理されるのである。

【0014】即ち、図4の選択回路1(3)からの選択信号により、スイッチ1(101)が選択され、必要な音声素片に対応する残差波形が、残差波形メモリ(8)から読み出され、残差波形バッファ1(102)に蓄えられる。「た\*べにき\*た」の例では、まず、「た」の残差波形が、読み出される。次に、選択回路2(9)からの選択信号により、スイッチ2(103)が選択され、適切なピッチの残差が選択され、残差波形バッファ2(104)に蓄えられる。例では、図2に示すように「た」のピッチは、400Hzと高いので、高音用の「た」の残差波形が選択されて、残差波形バッファ2(104)に蓄えられる。最後に、ピッチパターン生成部(7)で決定されたピッチになるように、残差波形のピッチ周期の変更がピッチ変更回路(105)で施される。例えば、高音用の「た」の残差波形が、380Hzの音声から抽出されたものであれば、ピッチパターン生成部(7)で決定された400Hzになるよう、20Hzだけピッチが低くなる(ピッチ周期が長くなる)ような変更が施される。

【0015】尚、ピッチ変更回路(105)で行われる残差波形のピッチの変更としては、例えば、ピッチを低くする時は、途中で零データを挿入してピッチ周期を長くし、ピッチを高くする時は、途中のデータを削除して

ピッチ周期を短くする「零詰め切り捨て法」が用いられる。

【0016】ピッチ変更を大幅に行うと、音質の劣化が生じるので、上記の残差駆動型規則合成装置の例では、高音用、中音用、低音用の3段階のピッチの異なる残差をあらかじめ残差波形メモリに蓄えておき、所望のピッチに近いピッチ周期の残差波形を用いることにより、ピッチの変更量が少なくすむように工夫している。

【0017】上述の如く、選択された音声素片に対応し、かつ、所望のピッチに変更された残差波形が、駆動音源として生成されるので、発音の自然性が高い合成音声を得られるのである。

【0018】以上に概説した残差駆動型音声合成装置によれば、その残差波形メモリ(8)に蓄えられているピッチの異なる残差波形は、従来は図5に示す方法で作成されるのが一般的であった。

【0019】即ち、例えば、3種類(3段階)のピッチの残差波形を作成する場合、図5(a)に示すように、高音の入力音声进行分析することにより、ピッチ周期の短い、高音用の残差波形を抽出するのである。また、図5(c)に示すように、低音の入力音声进行分析することにより、ピッチ周期の長い、低音用の残差波形を抽出する。中音のそれについても、図5(b)に示すように、中音の入力音声进行分析することにより、ピッチ周期の平均的な長さの、低音用の残差波形を抽出するのである。

【0020】

【発明が解決しようとする課題】前述した如く、従来の残差駆動型規則合成装置では、ピッチの異なる残差波形を利用して、駆動音源を生成する場合、音声素片として蓄えられているLPCやLSPなどのパラメータと、駆動音源として利用される残差波形が、異なる音声から分析して抽出されたものとなるため、蓄えられた残差波形を駆動音源とし、蓄えられた音声素片のパラメータを係数として合成フィルタに通しても原音声を再生することができず、生成される合成音声が悪化するという問題点が生じる。

【0021】また、各残差波形ごとに、対応する高さの音声から抽出した音声素片を利用すれば、劣化はなくなると考えられるが、この場合は各ピッチに対応して音声素片を複数個、蓄えておくことが必要となり、メモリ量が増大する。

【0022】

【課題を解決するための手段】本発明の残差駆動型音声合成装置は、音声合成フィルタの逆フィルタからなる音声分析フィルタの係数として、音声素片メモリの音声素片単位の音声パラメータを用い、この分析フィルタにそれぞれ異なるピッチ周期の音声を入力することにより得られるそれぞれ異なるピッチ周期の音声の残差信号の波形データを抽出し、この波形データを残差波形メモリに蓄えたものである。

## 【0023】

【作用】本発明の残差駆動型音声合成装置によれば、音声合成するために選択された音声素片に対して、どのような周期の残差波形が駆動音源として用いられても、選択され得る全ての残差信号の波形データが、実際の合成に供せられる音声パラメータそのものを係数とした逆フィルタにより抽出されたものであるため、ピッチの変更にかかわらず、音源とパラメータが同一音声から抽出したことになるため、より自然で高品質な合成音声を得られる。

## 【0024】

【実施例】図6は、本発明の残差駆動型音声合成装置に用いる残差信号の波形データの作成方法を示したものである。図6(a)(b)(c)の各 $H(z)$ は、音声合成フィルタの逆フィルタからなる音声分析フィルタの伝達特性をゼット変換の記述で表したものである。これらの図6(a)(b)(c)にそれぞれ示すように、高音、中音、低音の3段階のピッチの異なる音声に対して、音声素片として蓄えられるLPC(線形予測係数、編自己相関係数等)やLSPなどの音声パラメータを係数とする逆フィルタをかけることにより、それぞれ高音用、中音用、低音用の残差波形が生成される。

【0025】本発明の残差駆動型音声合成装置は、各音声素片に対し、図6に示したのと同様の方法で作成した残差波形を図3に示す残差波形メモリに蓄え、図1に示す構成で残差駆動型の規則合成を行うのである。

【0026】すなわち、発声すべき文字列が文字列バッファ(1)に入力されると、音韻記号列生成部(2)は入力された文字列を音韻記号列に変換する。例えば、「た\*べにき\*た」(但し、\*はアクセント位置を示す記号)という文字列が入力されると「tabeni k i t a」という音韻記号列に変換する。

【0027】次に、選択回路1(3)は、音韻記号列から必要な音声素片を順次、決定選択し、音声素片メモリ(4)に蓄えられた音声素片のうち必要な音声素片が、音声素片接続部(5)で接続される。音声素片の単位としては、CVC(子音+母音+子音)、あるいは、CV(子音+母音)とVC(母音+子音)を併用するものなども用いられるが、例として簡単のため、CV(子音+母音)すなわち音節を単位とするものを考えると、「t a、be、ni、ki、t a」が必要な音声素片として順次選択され接続される。

【0028】接続された音声素片は、音声パラメータとして音声パラメータバッファ(6)に蓄えられ、係数として合成フィルタ(11)に与えられる。

【0029】一方、アクセント位置等のイントネーションを表す記号も、発声すべき文字列とともに入力され、文字列バッファ(1)から、ピッチパターン生成部(7)に与えられると、ピッチパターン生成部(7)は、発声文全体のピッチ(音程)を決定する。例えば、

「た\*べに き\*た」という入力の場合、「た」と「き」にアクセントが存在するので、第5図に示すようなピッチパターンとなる。

【0030】また、残差波形メモリ(8)には、各音声素片に対応して、駆動音源として利用するための残差波形が蓄えられており、選択回路1(3)で決定選択された音声素片に対応して、必要な残差波形が選択される。例の場合、「t a、be、ni、ki、t a」の順で残差波形が選択される。

10 【0031】さらに、残差波形メモリ(8)には、各音声素片に対して、本発明で提案する方法で作成したピッチの異なる複数の残差波形が蓄えられており、ピッチパターン生成部(7)で生成されたピッチに応じて、選択回路2(9)が適切なピッチの残差波形を選択決定し、駆動音源生成部(10)に蓄える。最後に、選択された残差波形は、駆動音源生成部(10)で、ピッチパターン生成部(7)で生成されたピッチに一致する値にピッチ変更が施され、所望のピッチの残差波形が生成される。

20 【0032】このようにして生成された残差波形は、駆動音源として合成フィルタ(11)に入力され、合成フィルタ(11)で合成音声生成される。合成音声はD/A変換器(12)を経て、スピーカ(13)から出力される。

## 【0033】

【発明の効果】本発明の残差駆動型音声合成装置は、ピッチの変更にかかわらず、音源とパラメータが同一音声から抽出したことになるので、接続すべき音声素片に対応して選択された残差波形であれば、いずれのピッチの残差波形を駆動音源として用いても、原音声に近い高品質な合成音声を得ることができる。

## 【図面の簡単な説明】

【図1】残差駆動型規則合成装置の構成図、

【図2】ピッチパターンの模式図、

【図3】残差波形メモリの模式図、

【図4】駆動音源生成部の模式図、

【図5】従来の残差波形作成方法の解説図、

【図6】本発明の残差駆動型音声合成装置で用いる残差波形作成方法の解説図。

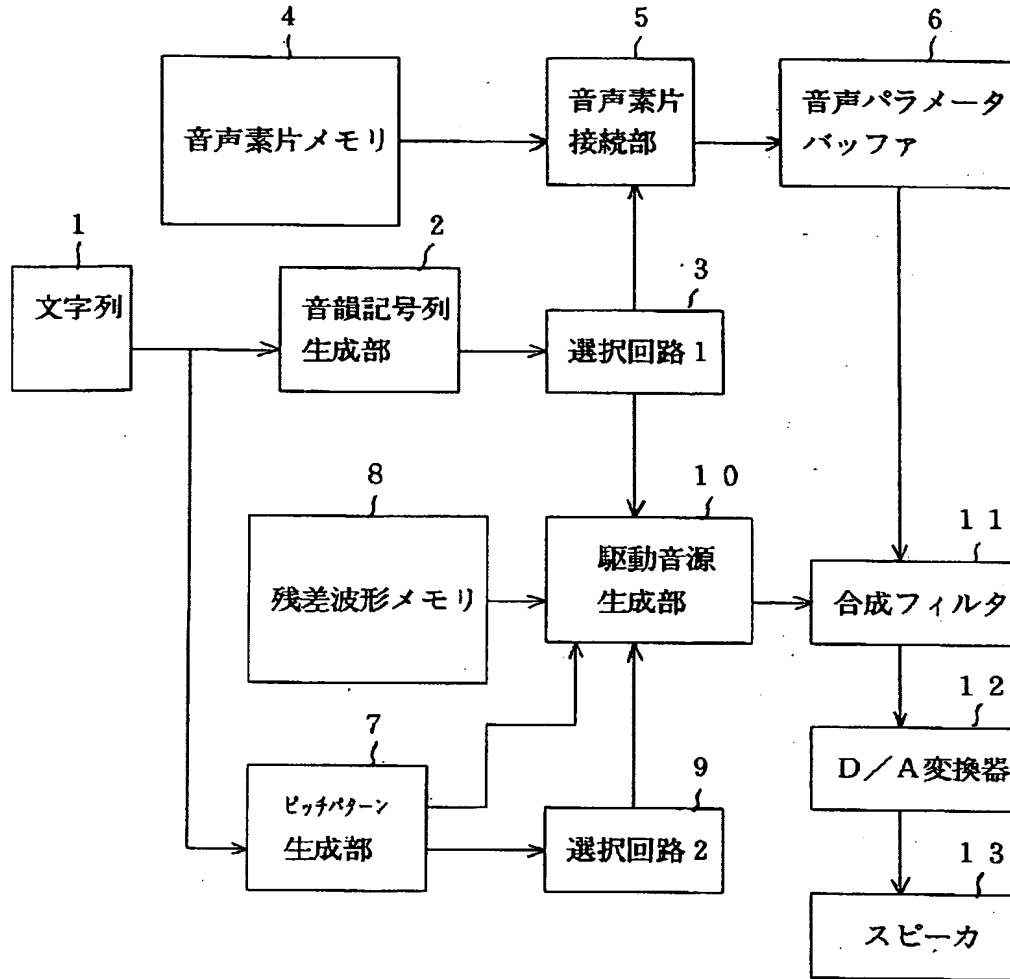
## 【符号の説明】

- (1)・・・文字列バッファ、
- (2)・・・音韻記号列生成部、
- (3)・・・選択回路1、
- (4)・・・音声素片メモリ、
- (5)・・・音声素片接続部、
- (6)・・・音声パラメータバッファ、
- (7)・・・ピッチパターン生成部、
- (8)・・・残差波形メモリ、
- (9)・・・選択回路2、
- (10)・・・駆動音源生成部、

( 1 1 ) ・ ・ 合成フィルタ、  
 ( 1 2 ) ・ ・ D / A 変換器、  
 ( 1 3 ) ・ ・ スピーカ、  
 ( 1 0 1 ) ・ スイッチ 1、

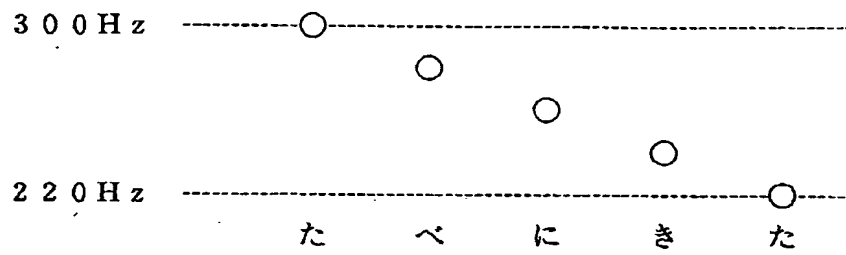
( 1 0 2 ) ・ 残差波形バッファ 1、  
 ( 1 0 3 ) ・ スイッチ 2、  
 ( 1 0 4 ) ・ 残差波形バッファ 2、  
 ( 1 0 5 ) ・ ピッチ変更回路

【図 1】

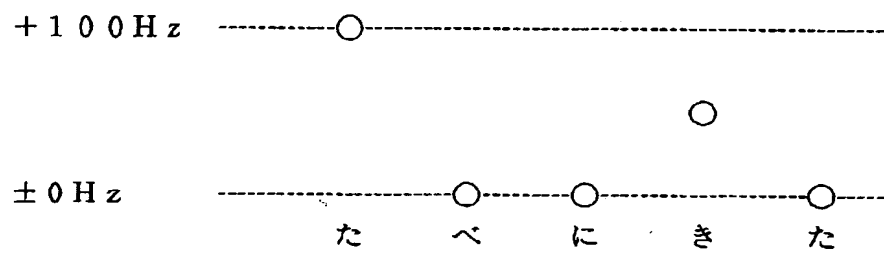


【図 2】

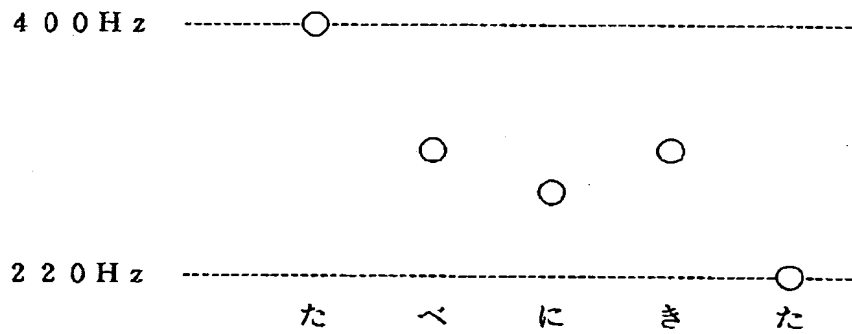
(イ) フレーズ成分



(ロ) アクセント成分



(ハ) ピッチパターン

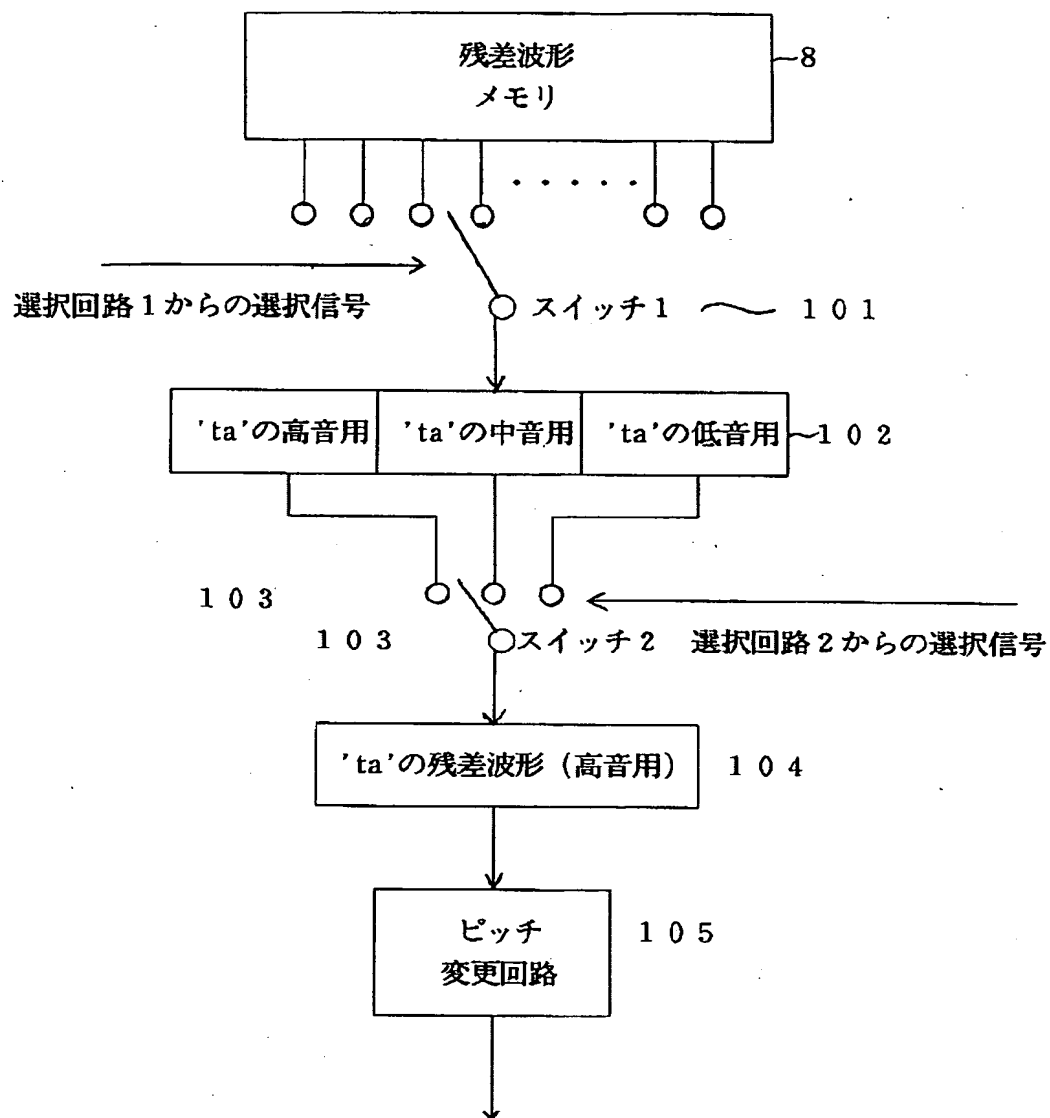


【図 3】

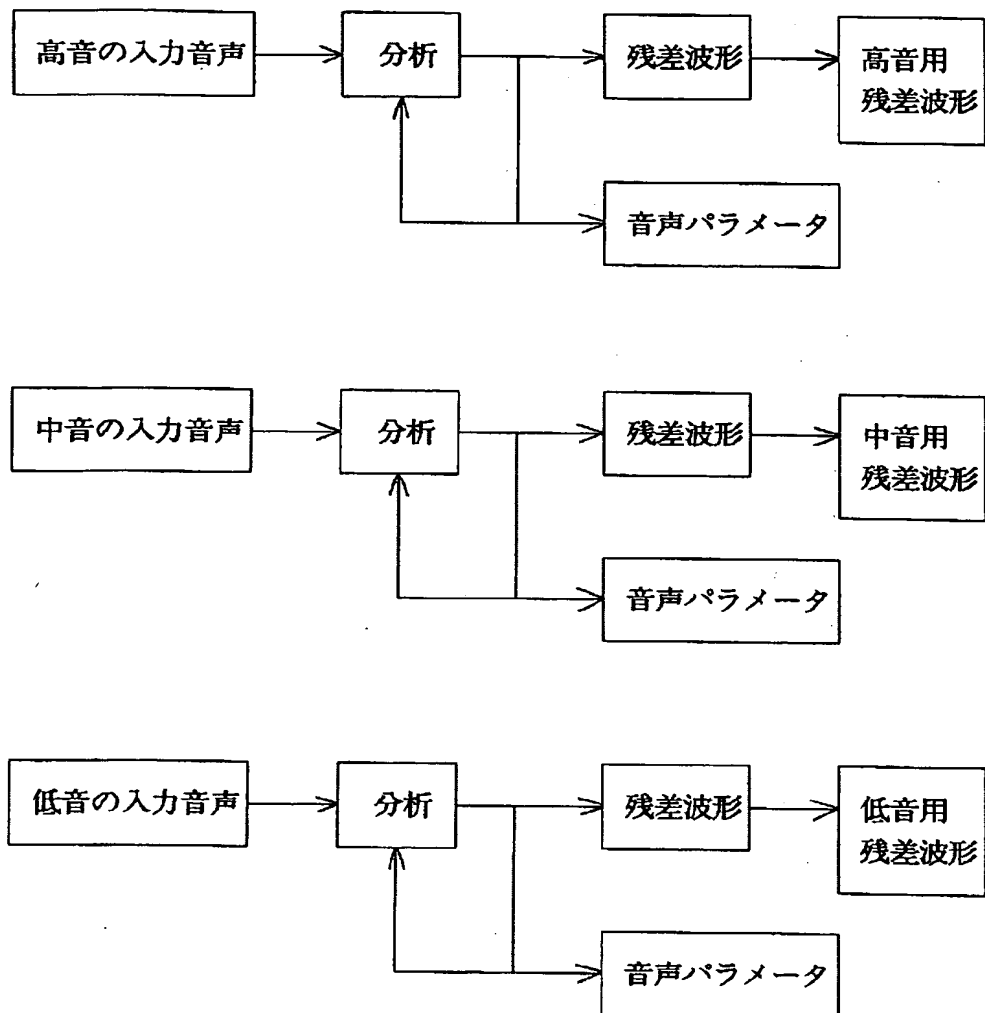
' a' の残差波形（高音用）
' a' の残差波形（中音用）
' a' の残差波形（低音用）
' i' の残差波形（高音用）
' i' の残差波形（中音用）
' i' の残差波形（低音用）
⋮
' t a' の残差波形（高音用）
' t a' の残差波形（中音用）
' t a' の残差波形（低音用）
⋮
⋮
⋮



【図 4】



【図 5】



【図 6】

